Dynamic Bit Encoding for Privacy Protection Against Correlation Attacks in RFID Backward Channel

Kazuya Sakai, *Student Member, IEEE,* Wei-Shinn Ku, *Member, IEEE,*
Roger Zimmermann, *Senior Member, IEEE,* and Min-Te Sun, *Member, IEEE*

Abstract—Nowadays Radio Frequency Identification (RFID) technologies are applied in many fields for a variety of applications. Though bringing great productivity gains, RFID systems may cause new security and privacy threats to individuals or organizations. Therefore, it is important to protect the security of RFID systems and the privacy of RFID tag owners. Unfortunately, none of the existing solutions provide a complete defense against eavesdroppers who could monitor the communication between RFID readers and tags and recover the contents of tags. Based on our research, we propose two novel RFID backward channel protection protocols, namely dynamic bit encoding and optimized dynamic bit encoding. Our schemes are able to achieve high anonymity with limited communication overhead. Our extensive simulations show that both proposed schemes provide much stronger backward channel protection than existing techniques. In addition, analytical models were created and validated through comparisons with simulation results.

Index Terms—Radio Frequency IDentification, RFID, privacy protection, bit encoding.

1 INTRODUCTION

Radio Frequency Identification (RFID) is an electronic tagging technology that allows objects to be automatically identified at a distance without a direct line-of-sight using an electromagnetic challenge-and-response exchange of data [29]. RFID systems smooth the way for various applications such as supply chain management [17], [24], transportation payment, animal identification, warehouse operations [5], and more. Though bringing great productivity gains, RFID systems may cause new security and privacy threats to individuals or organizations [9], [14], [22], [28]. In all the aforementioned applications, an RFID reader has to identify individual tag IDs within its reading range. When a reader queries for tag IDs, several tags may respond at the same time and this would result in signal collisions. Several anti-collision protocols have been proposed [8], [21], [30] and the binary tree-walking-based scheme is commonly used. With the binary tree-walking-based protocol, a reader broadcasts each bit of the singulated tag's ID over the long-range forward channel, and eavesdroppers who are within the signal range of the reader can monitor the process and recover the contents of every tag. In order to protect privacy, a number of techniques [30], [31] have been designed to safeguard the forward channel in RFID systems.

- *Kazuya Sakai is with the Department of Computer Science and Engineering, Ohio State University, Columbus, OH, USA 43210.*
  E-mail: sakai.16@buckeyemail.osu.edu
- *Wei-Shinn Ku is with the Department of Computer Science and Software Engineering, Auburn University, Auburn, AL, USA 36849.*
  E-mail: weishinn@auburn.edu
- *Roger Zimmermann is with the Department of Computer Science, National University of Singapore, Singapore 117417.*
  E-mail: rogerz@comp.nus.edu.sg
- *Min-Te Sun is with the Department of Computer Science and Information Engineering, National Central University, No. 300, Jhoogda Rd. Jhongli, Taoyuan 320, Taiwan.*
  E-mail: msun@csie.ncu.edu.tw

Manuscript received 16 Apr. 2010; revised 28 Apr. 2011; accepted 7 Oct. 2011.

The backward channel – transmitting data from a tag to the reader – has much lower signal energy than the forward channel and is therefore more difficult to observe. However, an attacker could be in close proximity of a tag and s/he may listen to the communication of the backward channel. This will also threaten the privacy of the tag owner. For example, assume a retail store installs an RFID-based smart shelf system for managing RFID-tagged products. In such a setting, an attacker (e.g., a corporate spy from a competitor store) could collect the entire store's inventory and sales data by pretending to be a customer, if the smart shelf system has no backward channel protection against eavesdropping. Therefore, it is equally important to protect the backward channel. The solution proposed in [3] relies on the reader to transmit a mask bit string concurrently with a tag sending out its identifier through the backward channel. This will result in signal collisions and partially obstructed readings. Since the reader knows the mask bit string, the RFID tag can be successfully reconstructed. Nevertheless, this method suffers from *the same bit problem*, in which some bits of the tag ID could still be disclosed. An improved solution based on [3] is presented in [18] by encoding each source bit of the tag into a fixed length n-bit string to alleviate the same bit problem. This solution is still vulnerable if an attacker has knowledge of the n-bit string length and can interrogate a tag repeatedly (detailed in the Motivations subsection of Section 3). Consequently, we need more advanced techniques that provide much stronger RFID backward channel privacy preserving capabilities than existing solutions.

In this paper we put forth two novel RFID backward channel protection protocols. The contributions of this research are as follows.

- We propose a bit encoding scheme, namely Dynamic Bit Encoding (DBE) for privacy protection in RFID backward channels. DBE encodes the i-th source bit based on all the preceding $(i-1)$ source bits, which makes it very difficult to crack the original ID.
- We further improve the degree of security of DBE and design an Optimized Dynamic Bit Encoding (ODBE) scheme by dynamically changing the maximum codeword length for each source bit.
- Analytical models of guessing attacks [18], anonymity against generated encoded ID and correlation attacks are created and validated.
- Analyses for the communication overhead and time complexity incurred by DBE and ODBE are conducted.
- We evaluate our proposed techniques through extensive simulations. The results show that our DBE and ODBE schemes provide a much more robust backward channel protection than previous techniques with the same communication overhead.

The rest of this paper is organized as follows. Section 2 surveys related works. The DBE and ODBE schemes are introduced in Section 3. Analytical models are provided in Section 4. Section 5 presents the analysis for control overhead. The experimental validations of our protocols are presented in Section 6. Section 7 concludes the paper with a discussion of future work.

2 Related Work

In this section we review RFID singulation protocols and previous work related to our approach of RFID system backward channel protection.

2.1 RFID Singulation Protocols

In RFID systems, a reader has to recognize individual tag IDs in its reading region. However, collisions may happen when several tags respond simultaneously to the reader query. Therefore, we need anti-collision singulation schemes for a reader to effectively identify tags in its proximity. Current singulation protocols can be roughly categorized into Aloha scheme based protocols and tree-walking scheme based protocols.

In Aloha-based protocols [8], [20] – named after an early wireless network protocol developed at the University of Hawaii – a reader sends a query frame and each tag randomly chooses a time slot to send its ID information. If more than two tags select the same slot, collisions occur. The colliding tags have to choose another slot to send a response. In addition, the reader can adjust the frame size according to the number of collisions in the previous frame. Although Aloha-based protocols avoid collisions to identify tags, a specific tag may not be identified for a long time – this scenario is also called the tag starvation problem.

In tree-walking-based protocols [21], [25], [31], a reader traverses a binary tag tree, which organizes the entire ID space of tags and each tag ID is mapped to a leaf node in depth-first or breadth-first order. For singulation, a reader broadcasts a query to all tags in the vicinity for the next bit of their ID. On receiving a query, a tag responds if its ID matches the prefix of the bit string in the query. If more than one tag responds, the reader will be able to detect the collision. Afterward the reader will broadcast a bit indicating whether tags who replied with a 0 or those who replied with a 1 should continue. By applying this mechanism, all tags in the interrogation area will be identified. While tree-walking-based protocols may incur a long singulation delay, they do not suffer from the tag starvation problem that occurs with Aloha-based protocols.

Both Aloha and tree-walking-based protocols can be implemented with simple methods such as randomly selecting a time slot and matching query bits with tag IDs to identify low-cost RFID tags. However, all the aforementioned singulation protocols do not support privacy protection of the communication between readers and tags.

2.2 Encryption-based Authentication in RFID

As passive tags are computationally weak devices, they cannot perform traditional cryptographic techniques such as symmetric key and public/private key operations. This enforces a number of authentication solutions for RFID systems to use low-cost cryptographic operations [13]. The basic idea of encryption-based authentication techniques employed in RFID systems works as follow: Assume a reader and a tag share a common secret key, say key K, which is a m-bit random string. When the tag sends its real ID, it calculates $ID \oplus K$, where $\oplus$ is the XOR operation. The reader decodes the cipher text with the key, and it successfully reads the tag without being eavesdropped. This leads many studies to emphasize on secure key exchange and distributions [7], [15], [19]. However, key-searching and key-updating operations are usually expensive in large-scale RFID systems. In this paper, we focus on tag authentication without a shared secret key between a reader and a tag. The most related works to this paper are reviewed in the following subsection.

2.3 Privacy Protection in RFID Singulation

Since every bit of every singulated tag is broadcast by the reader on the forward channel, attackers could monitor these transmissions from a significant distance and recover the contents of every tag. Weis [30] proposed two secure tree-walking protocols for protecting the forward channel from eavesdroppers. In the blinded tree-walking algorithm [31], when there is no collision in a certain bit position, instead of specifying which portion of the tag population should proceed, readers send the query signal for the next ID bit directly, hence not all the bits are transmitted on the forward channel. In the randomized tree-walking algorithm, each tag has two IDs – a real one and a pseudo-ID allocated by manufacturers or generated by the tag itself. Readers singulate with pseudo-ID values and tags respond with their real IDs over the backward channel.

Eavesdroppers may appear near an RFID tag and clandestinely listen on the backward channel, i.e., the signals sent from a tag to a reader, which leads to privacy threats. Choi

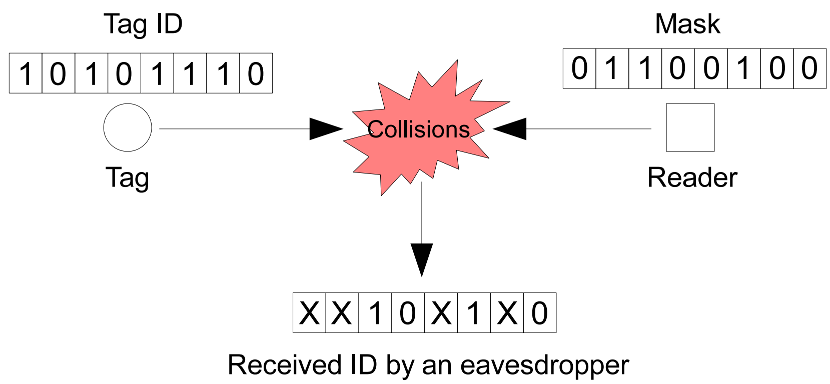

Fig. 1. Privacy masking.

and Roh [3] designed the privacy masking protocol to protect the transmission of tag IDs by making the reader transmit a mask at the same time as the tags transmit their data. As shown in Figure 1, collisions occur between the mask and the response tag ID, and consequently eavesdroppers can only obtain IDs in which a subset of bits are unidentifiable. Only when a mask has the same bit sequence as a tag ID (i.e., the same bit problem), the real ID of the tag will be disclosed. Therefore, this solution is able to protect tag identifiers from adversaries when a legitimate reader is singulating tags. However, the mechanism does not provide any protection against unauthorized readers which can also singulate tags.

Lim et al. [18] introduce a randomized bit encoding scheme (RBE) which is able to alleviate the same bit problem of the technique in [3]. In RBE, each bit of a real tag ID is encoded into an n-bit string and a tag sends its encoded ID to a reader under the privacy masking protocol. Afterward the reader decodes each n-bit string. If the Hamming weight of the n-bit string is even, the corresponding source bit is 0. Otherwise the source bit is 1. By employing this method, even if some bits of the n-bit string did not collide, it would be difficult for an attacker to recover the original ID. In addition to the encoding scheme, Lim et al. propose a new system architecture against unauthorized readers, which is illustrated in Figure 2. Instead of the RF reader, a separate trusted device, called Trusted Masking Device (TMD), is responsible for sending the mask. This architecture protects the privacy of tags from unauthorized readers and is extensively studied in [2], [16], [23]. Besides, a similar model which focuses on the combination of simultaneously received electromagnetic signals has been employed in the notable physical layer network coding (PNC) scheme [1], [32], allowing a multi-hop wireless network to achieve the maximum possible information flow.

3 System Design

In this paper, we adopt the architecture in [18] and assume the deployment of TMDs. In addition, we assume that each tag takes part in the singulation process with a randomly generated pseudo ID. Based on these assumptions, we propose two novel bit encoding schemes for the purpose of privacy masking to protect the backward channel, namely Dynamic Bit Encoding (DBE) and Optimized Dynamic Bit Encoding (ODBE). The notations used in this paper are listed in Table 1.

TABLE 1
Definition of notations.

Symbol	Meaning
C	The constant length of an encoded ID
n	The length of a codeword
b_i	The i-th source bit
n_i	The number of bits for encoding the i-th source bit
N_{max}	The maximum value of n
N_i	N_{max} for the i-th bit
$E(i,n)$	The codeword of b with length n
$F(key)$	A hash function
P	The correct guess probability
l	The length of a tag ID
l_c	The number of unreadable bits
pr_i	The probability of the i-th bit is identifiable
ϕ	The set of all possible IDs
ϕ'	The anonymous set
H_ϕ	The entropy of the system ϕ
$d_{\phi'}$	The degree of anonymity
R	The number of readable bits
I_{rate}	Information rate to measure the effectiveness of coding schemes
S_t	The set of possible real tag IDs, i.e., $\{0,1\}^l$
l_{tx}	The length of transmitted IDs (a real or pseudo ID)
l_p	The length of prefix in tree-walking singulation
s	The number of slots in Aloha-based singulation
T	The number of RF tags in the system

3.1 Motivations

To protect the privacy of an RFID tag, a protocol should incur a low probability that the original ID is identified (in the rest of the manuscript we will call this value the *correct guess probability* and denote it with P). Ideally, none of the bits in the original ID should be readable by attackers, i.e., all bits in

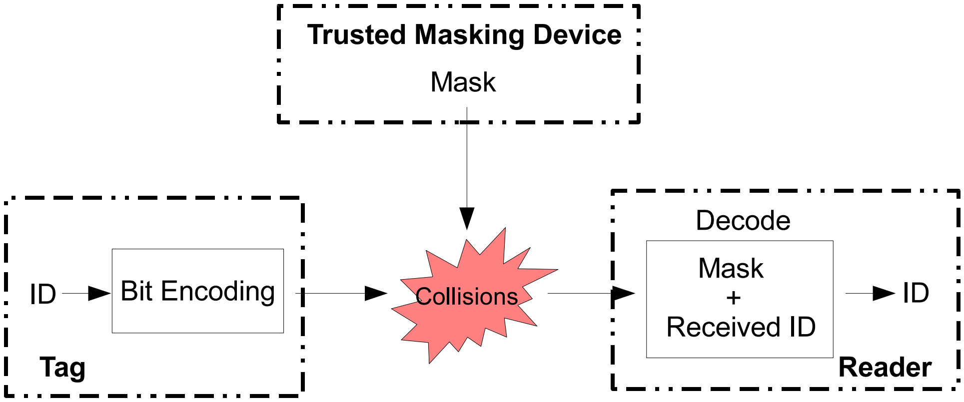

Fig. 2. Randomized bit encoding with a trusted masking device.

the ID or encoded ID should collide with the mask. In such a case the attackers achieve the lowest correct guess probability. When a bit is not readable, the correct guess for the bit has a chance of 50%. Let l be the ID length and l_c be the number of unreadable bits. In the best case, $l = l_c$ and $P = (\frac{1}{2})^l$. On the other hand, when a mask has the same bit sequence as the ID or the encoded ID, the correct guess probability is 100% (i.e., the same bit problem). Thus, we can establish upper and lower bounds for P: $(\frac{1}{2})^l \leq P \leq 1.0$.

In [3], the probability that each bit collides is 0.5. When a bit collides, the probability that an attacker can identify the source bit is 0.5. On the other hand, if a bit does not collide, the probability is 1. Therefore, the correct guess for the privacy masking protocol without encoding is as follows.

$$P = \left\{\frac{1}{2} \cdot 1 + \frac{1}{2} \cdot \frac{1}{2}\right\}^l = \left(\frac{3}{4}\right)^l \tag{1}$$

For the randomized bit encoding (RBE) scheme [18], the probability that all bits in a codeword do not collide with the mask bit string is $\frac{1}{2^n}$ with the optimal n-bit encoding. The probability that a bit is identifiable is $\frac{1}{2^n} + (1 - \frac{1}{2^n}) \cdot \frac{1}{2}$. Therefore, the correct guess probability of the RBE scheme is:

$$P(n) = \left\{\frac{1}{2^n} + (1 - \frac{1}{2^n}) \cdot \frac{1}{2}\right\}^l, \quad (n \geq 1) \tag{2}$$

Both the privacy masking and RBE protocols provide protection for the backward channel against guessing attacks to identify the original ID. However, after a number of interrogation cycles, attackers can obtain the original ID from collected source bits – this is called *the correlation attack of encoded IDs* (termed correlation attack for the rest of this paper). As demonstrated in Figure 3, an attacker is able to receive encoded IDs by interrogating the tag repeatedly. The original ID '101' can be identified by decoding the readable codewords of collided IDs received in all the previous cycles. Unfortunately, the two aforementioned protocols are vulnerable to the correlation attack (the analysis is provided in Section 4). To deal with the correlation attack, a stronger tag ID encoding mechanism must be designed.

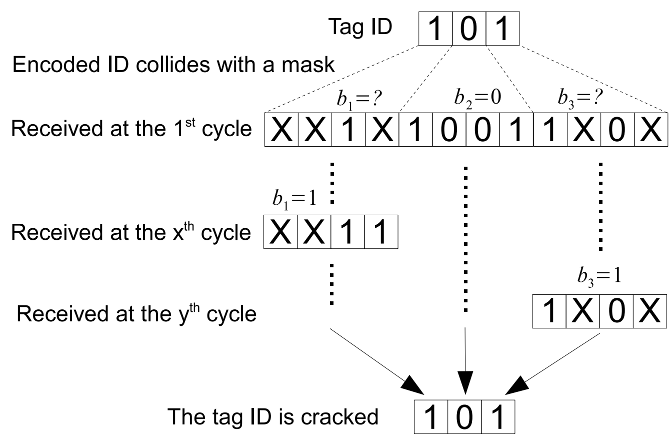

Fig. 3. The correlation attack of an encoded ID.

3.2 Dynamic Bit Encoding Scheme

In the randomized bit encoding scheme [18], the encoding of a source bit does not depend on any other bits in the ID. In other words, a source bit can be identified independently. Our fundamental idea is that the codeword to encode the i-th source bit should be determined based on all the preceding $(i-1)$ source bits. By employing this design, the i-th bit of the original ID cannot be identified until all the previous $(i - 1)$ bits are identified.

In dynamic bit encoding, the source bits in the original ID are encoded by codewords of variable length n where $n \in \{1, 2, ..., N_{max}\}$. Let b_i be the i-th bit in the original ID, and n_i be the value of n for the i-th bit. The first bit b_1 is always encoded by a codeword of length N_{max}, and the codeword of b_i is denoted by $E(b_i, n_i)$ where $E(b_i, n_i) \in \{0, 1\}^{n_i}$. If the Hamming weight of the first i bits is odd, $E(b_i, n_i)$ returns a bit string whose Hamming weight is odd. Otherwise, it returns a bit string whose Hamming weight is even. For example, if the Hamming weight of the first i bits is odd, $E(b_i, 3) \in \{001, 010, 100, 111\}$, and $E(b_i, 3) \in \{000, 011, 110, 101\}$, if the Hamming weight is even. A tag randomly picks one of the codewords in the set. By applying this design, if one of the codeword bits collides with a mask bit, the corresponding source bit is unidentifiable. In addition, the codeword length of the i-th bit (except b_1) is determined by the result of a hash function, $F(key) \in \{1, 2, ..., N_{max}\}$. The key needs to be associated with the codeword of the previous bit to achieve our design goal, and the length of the codeword for the i-th bit is $F(E(b_{i-1}, n_{i-1}))$. Most well-known deterministic hash methods can be used for this system as long as they return values uniformly in the range of 1 to N_{max}. Consequently, the codeword length of b_2 is $F(E(b_1, n_1))$ bits. By repeating these steps, all the bits in the original ID can be encoded.

After encoding all the source bits, the number of bits of the corresponding encoded ID is $\sum_{k=1}^{l} n_k$. We concatenate an extra $l \cdot N_{max} - \sum_{k=1}^{l} n_k$ bits to the end of the encoded ID to pad it to a constant length C. Accordingly, the overhead of this protocol is $l \cdot N_{max} - l$. There are two main reasons to append extra bits to the encoded ID to make it $l \cdot N_{max}$ bits long. The first is to improve privacy protection. For example, assume that the length of an encoded ID is $N_{max} + l - 1$ (i.e., the first bit is encoded by N_{max} bits and other bits are encoded with 1 bit) and the original ID is 1011 with $N_{max} = 3$. The encoded ID could be '100011'. An attacker with background knowledge is able to infer that the first bit is encoded by 3 bits, and other bits are encoded by 1 bit from the length of the encoded ID. The second reason is that it simplifies TMD, since it only needs to know the length of the ID and N_{max} to create a mask bit string.

Note that in DBE, encoded IDs generated from different tag IDs may collide. However, such situations are rare, and in reality the singulation with DBE should not suffer from encoded ID collisions (see the proof in Appendix A).

3.3 Examples of the Dynamic Bit Encoding Scheme

In this subsection, examples of DBE are presented. Consider Figure 4 where the original tag ID is '1011', $l = 4$, and $N_{max} = 3$. A hash function, $F(key) = key \bmod N_{max} + 1$ is used. The first bit $b_1 = 1$ is encoded by 3 bits. Assume '111' is chosen as the codeword of $E(1, 3)$ and $F(111) = 7 \bmod 3 + 1 = 2$. Thus, the second bit $b_2 = 0$ is encoded by 2 bits. Since the Hamming weight of the first 2 bits, i.e., '10' , is odd, the Hamming weight of the codeword of b_2 has to be odd. Further following the encoding scheme, the encoding result could be '11101001'. To make the length of the encoded ID equal to C, extra random bits are appended. The tag sends

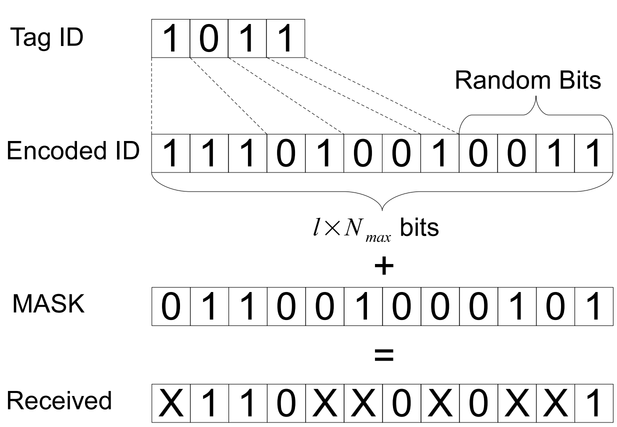

Fig. 4. An example of DBE.

the encoded ID '111010010011' to the reader with the privacy masking mechanism. Suppose the mask is '011001000101', then an eavesdropper receives 'X110XX0X0XX1' after the collision. Since the first bit is unrecoverable, the eavesdropper cannot identify any bit of the original ID.

The reader also receives the collided bit string. In contrast to the attacker, as the reader has the knowledge of the mask, it can recover the encoded ID. From the encoded ID, the reader obtains the original ID as follows. Because the reader knows $N_{max} = 3$ (i.e., $n_1 = 3$), it can find $E(b_1, 3) = 111$. Since the Hamming weight of '111' is odd, the reader obtains $b_1 = 1$. By computing the hash function, the reader acquires $n_2 = 2$. Because the Hamming weight of $E(b_2, 2) = 01$ is odd, the Hamming weight of the first 2 bits (i.e., '1X') has to be odd. As we already know $b_1 = 1$, the reader derives $b_2 = 0$. By repeating this process, the reader can retrieve the original ID of the tag.

3.4 Optimized Dynamic Bit Encoding

To further improve the degree of security for RFID backward channel protection, an Optimized Dynamic Bit Encoding (ODBE) scheme was designed based on DBE. In ODBE, the value of N_{max} is dynamically changed for each source bit. Let N_i be the value of N_{max} for the i-th bit. With a randomly generated value for n, the length of the first codeword is decided by $n_1 = n$. Then the length of the i-th codeword is decided by $F(key) = key \mod N_i + 1$, where $N_i = n \cdot i - \sum_{k=1}^{i-1} n_k$. For the last source bit, its codeword should use up all the remaining bits l_l in the encoded ID after $(l-1)$ source bits in the original ID were encoded. The value of l_l can be obtained as follows:

$$l_l = n \cdot l - \sum_{k=1}^{l-1} n_k \tag{3}$$

The last bit is encoded by a codeword with length $n_l = l_l$. Consequently, the length of the encoded ID is always $n \cdot l$ bits. An example of ODBE is shown in Figure 5 with $n_1 = 3$. ODBE provides a higher degree of security than DBE, which is postulated in the following lemma:

Lemma 1 *When the randomly generated first codeword length n in ODBE is equal to N_{max} in DBE, ODBE results either in lower or equal correct guess probability than DBE.*

Proof: To prove the above claim, we show that the average value of codeword length for ODBE, $\overline{n}$ is greater than or equal to that of DBE. Let $\overline{n_i}$ be the average value of codeword length for the i-th bit in ODBE defined by:

$$\overline{n_i} = \begin{cases} n & \text{(if } i = 1) \\ \frac{2n - \overline{n_{i-1}}}{2} & \text{(if } i > 1) \end{cases} \tag{4}$$

The average value of codeword length of ODBE is:

$$\overline{n} = \begin{cases} \frac{1}{i}\sum_{k=1}^{i} \overline{n_i} & \text{(if } i \text{ is small)} \\ \frac{2n - \overline{n}}{2} = \frac{2}{3}n & \text{(if } i \text{ is sufficiently large)} \end{cases} \tag{5}$$

Note that $\frac{2}{3}n$ is the lower bound of $\overline{n}$ in Equation 5.

On the other hand, the average value of codeword length for the i-th bit in DBE is:

$$\sum_{k=1}^{N_{max}} k = \frac{1}{2}N_{max} = \frac{1}{2}n \tag{6}$$

Therefore, the average value of codeword length for ODBE is larger than that for DBE. This concludes the proof. □

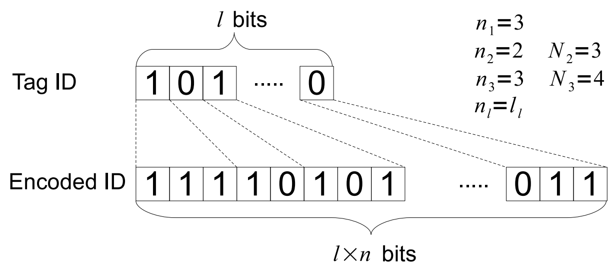

Fig. 5. An example of ODBE.

3.5 Implementation Considerations

The privacy masking scheme proposed by Choi and Roh [3], [4] is designed for a tree-walking-based protocol under the assumption that the bit-timing between a tag ID and a mask is properly synchronized. The bit timing synchronization required by this scheme can be realized by the adaptive demodulation of RFID receivers [11]. The fast synchronization in [11] is achieved by demodulating the burst-receiving data and decoding with fewer frequency deviation.

Privacy masking techniques can be implemented not only in tree-walking-based protocols, but also in Aloha-based protocols [8], [20]. In an Aloha-based singulation process, a tag randomly selects a time slot and sends a random number. It then waits for an acknowledgement from the reader. With this handshake, the reader and the tag agree on a time slot during which the pseudo ID is transmitted. Thus, all bits of the pseudo ID and the mask can be synchronized with each other.

The proposed DBE and ODBE protocols do not assume the underlying singulation scheme. In other words, our protocols can be applied to both tree-walking and Aloha-based singulation protocols.

3.6 Privacy Protection with CRC

Cyclic Redundancy Codes (CRC) are commonly used to correct unreadable bits due to communication errors. With CRC, if some of the bits in a data frame are corrupted, the whole frame can be recovered. The use of CRC improves the performance and robustness of communications. However, bit error corrections may lead to lower security. Backward channel protection with CRC has been extensively studied in [4]. The numerical results show that the probability of successfully eavesdropping increases when CRC is employed. Furthermore, CRC introduces extra control and communication overhead. Therefore, we do not consider the application of CRC in this paper.

4 Privacy Protection Analysis

In this section, analytical models of privacy protection are created for both DBE and ODBE.

4.1 Correct Guess Probability

First we analyze the *correct guess*, which is the probability of an attacker to guess the original ID. For the DBE scheme, the codeword length is among $\{1, 2, ..., N_{max}\}$ and the average codeword length is defined by Equation 6.

Note that from Equation 6, DBE has a higher codeword correct guess probability than RBE, since the probability is higher than $\frac{1}{2^n}$ when $n = N_{max}$. However, DBE is less vulnerable to the guessing attack because an attacker has to identify the preceding $(i-1)$ bits to recover the i-th bit. If an attacker knows N_{max}, given a DBE-encoded pseudo ID, the probability that exactly the first i bits of the original ID are identifiable (denoted as pr_i) is:

$$pr_i(N_{max}) = \begin{cases} 1 - \frac{1}{2^{N_{max}}} & (\text{if } i = 0) \\ \frac{1}{2^{N_{max}}} \cdot (\frac{1}{2^{N_{max}/2}})^{i-1} \cdot \{1 - \frac{1}{2^{N_{max}/2}}\} & (\text{if } i \geq 1) \end{cases} \tag{7}$$

When i number of source bits are disclosed to the attacker, the remaining $(l-i)$ source bits can be guessed with a 50% chance of correctness for each bit. The correct guess of the original ID when i source bits are identified can be denoted by:

$$pr_i(N_{max}) \cdot (\frac{1}{2})^{l-i} \tag{8}$$

Consequently, by combining Equations 7 and 8 the correct guess probability of the original ID is:

$$\begin{aligned} P(N_{max}) \quad = \quad & (1 - \frac{1}{2^{N_{max}}}) \cdot (\frac{1}{2})^l + \sum_{i=1}^{l} (\frac{1}{2^{N_{max}}}) \cdot \\ & (\frac{1}{2^{N_{max}/2}})^{i-1} \cdot \{1 - \frac{1}{2^{N_{max}/2}}\} \cdot (\frac{1}{2})^{l-i} \end{aligned} \tag{9}$$

For ODBE, pr_i is defined by Equation 10,

$$pr_i(n) = \begin{cases} 1 - \frac{1}{2^n} & (\text{if } i = 0) \\ (\frac{1}{2^n}) \cdot (\frac{1}{2^{\overline{n}}})^{i-1} \cdot \{1 - \frac{1}{2^{\overline{n}}}\} & (\text{if } i \geq 1) \end{cases} \tag{10}$$

Note that the average value of the codeword length for ODBE $\overline{n}$ is defined by Equation 5.

Therefore, the correct guess probability can be obtained as shown in Equation 11.

$$P(n) = (1 - \frac{1}{2^n}) \cdot (\frac{1}{2})^l + \sum_{i=1}^{l} (\frac{1}{2^n}) \cdot (\frac{1}{2^{\overline{n}}})^{i-1} \cdot \{1 - \frac{1}{2^{\overline{n}}}\} \cdot (\frac{1}{2})^{l-i} \tag{11}$$

Figure 6 demonstrates the analytical results of the correct guess probability as a function of different n values; here $n = N_{max}$. The correct guess is lower-bounded when the all bits are not readable, i.e., $(\frac{1}{2})^l$. As the value of n increases, the correct guess probability decreases for all the solutions. When

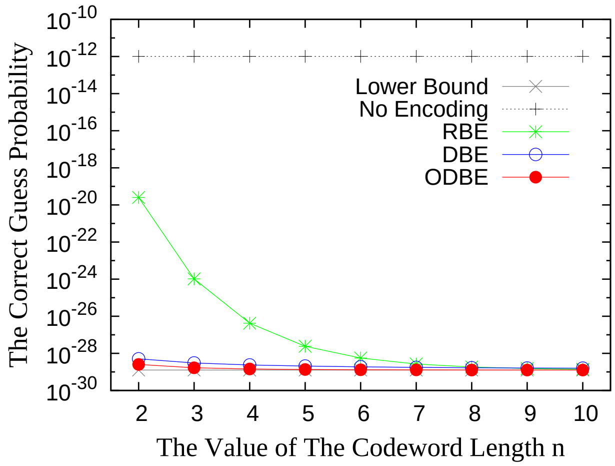

Fig. 6. The correct guess probability for different schemes.

$n \geq 7$, the correct guess probabilities of DBE, ODBE and RBE are all very close to the lower bound. When n is between 2 and 6, DBE and ODBE achieve a much reduced correct guess probability compared with RBE. Even with small n values, the correct guess probabilities of DBE and ODBE are very close to the lower bound and ODBE is slightly better than DBE. This analysis illustrates that both DBE and ODBE are less vulnerable to the guessing attack than RBE.

4.2 The Anonymity of Encoded IDs

In this subsection we introduce the *degree of anonymity* which is a privacy protection metric for encoded IDs [6], [26]. Anonymity is the state of not being identifiable within an anonymous set, and an anonymous set is the set of all possible IDs with similar characteristics as the original ID. For example, if an eavesdropper receives a 4-bit ID, '01XX', the corresponding anonymous set includes $\{0100, 0101, 0110, 0111\}$. The degree of anonymity for RFID systems can be defined by the entropy-based metric proposed in [12]. Consider a set ϕ of all possible IDs ($|\phi| = 2^l$) and a probability p_i of an ID being the original. The entropy of this system $H(\phi)$ is defined by:

$$H(\phi) = -\sum_{i \in \phi} p_i \log_2(p_i) \tag{12}$$

When all the bits are not readable, the system has the maximum entropy denoted by H_{max}. As guessing attacks generate probabilities with a uniform distribution, each element in the anonymous set has the same probability, i.e., $\forall i, p_i = (\frac{1}{2^{l_c}})$. Ideally, $l_c = l$. Therefore,

$$H_{max} = -\sum_{i \in \phi} \frac{1}{2^l} \log_2(\frac{1}{2^l}) = \log_2(2^l) = l \tag{13}$$

The degree of anonymity can be defined as:

$$\frac{H(\phi)}{H_{max}} \tag{14}$$

Let ϕ' be the possible ID set ($|\phi'| = 2^{l_c}$), when l_c number of source bits are not readable. Given ϕ', the degree of anonymity for a generated encoded ID, denoted as $d_{\phi'}$, is defined by:

$$d_{\phi'} = -\sum_{i \in \phi'} \frac{1}{2^{l_c}} \log_2(\frac{1}{2^{l_c}}) \cdot \frac{1}{l} = \frac{l_c}{l} \tag{15}$$

For example, consider a bit string of 'XX10X1X0' received after the encoded ID collided with the mask. Clearly, $l = 8$ and $l_c = 4$. Accordingly, $d_{\phi'} = 0.5$.

4.3 Correlation Attacks

To the best of our knowledge, no previous research has discussed correlation attacks in RFID system backward channels. The analytical models of correlation attack for our DBE and ODBE schemes as well as the privacy masking and RBE techniques are presented here.

For the privacy masking protocol, the probability that a source bit does not collide with the mask is 0.5. The number of bits which are identified after the t-th interrogation cycle, denoted as $R(t)$, is formalized by the following equations.

$$\begin{aligned} R(1) &= l \cdot \frac{1}{2} && (16) \\ R(2) &= R(1) + (l - R(1)) \cdot \frac{1}{2} && (17) \\ R(3) &= R(2) + (l - R(2)) \cdot \frac{1}{2} && (18) \\ R(t) &= R(t-1) + (l - R(t-1)) \cdot \frac{1}{2} && (19) \\ &= l \cdot (1 - \frac{1}{2^t}) && (20) \end{aligned}$$

For the randomized bit encoding protocol, the fraction $\frac{1}{2}$ in the above equations is replaced by $\frac{1}{2^n}$ and we can derive the following equation:

$$\begin{aligned} R(1) &= l \cdot \frac{1}{2^n} && (21) \\ R(t) &= R(t-1) + (l - R(t-1)) \cdot \frac{1}{2^n} && (22) \\ &= l \cdot \{1 - (1 - \frac{1}{2^n})^t\} && (23) \end{aligned}$$

For DBE and ODBE, the $(i-1)$-th bit needs to be identified to recover the i-th bit. Recall the symbol pr_i defined in Equations 7 and 10 for DBE and ODBE, respectively, which is the probability that the i-th bit is identifiable. Hence,

$$R(1) = \sum_{k=1}^{l} k \cdot pr_k \tag{24}$$

and, $R(t)$ is formalized by:

$$R(2) = R(1) + \sum_{k=1}^{l-R(1)} k \cdot pr_k \quad (25)$$

$$R(3) = R(2) + \sum_{k=1}^{l-R(2)} k \cdot pr_k \quad (26)$$

$$R(t) = R(t-1) + \sum_{k=1}^{l-R(t-1)} k \cdot pr_k \quad (27)$$

Equation 27 is too complicated to solve, and thus we simplify the equation. Since pr_k is very small and can be ignored when the value of $l - R(t-1)$ is large, we can approximate:

$$R(t) = R(t-1) + \sum_{k=1}^{l} k \cdot pr_k \quad (28)$$

$$= t \cdot \sum_{k=1}^{l} k \cdot pr_k \quad (29)$$

Thus, based on Equation 15, the tag anonymity at t for each protocol, denoted as $d_{\phi'}(t)$, is:

$$d_{\phi'}(t) = \frac{l - R(t)}{l} \quad (30)$$

We compare our analytical results with simulation results in the following section.

5 Analysis for Control Overhead

In this section, we study the communication overhead and the expected time for tag singulation of the proposed schemes.

5.1 Communication Overhead

To measure the effectiveness of coding schemes, the information rate [27] is generally employed, which is defined by Equation 31.

$$I_{rate} = \frac{1}{l_{tx}} \log_2 |S_t| \quad (31)$$

where l_{tx} is the length of transmitted IDs (a real tag ID or pseudo ID) over backward channel, and S_t is the set of possible real tag IDs. The base of log is 2, as tag IDs are represented by the binary system. Without loss of generality, $0 \leq I_{rate} \leq 1$. Clearly, for privacy masking without encoding [3], $l_{tx} = l$, $|S_t| = 2^l$, and hence $I_{rate} = 1$, which implies no redundancy. In contrary, RBE [18], DBE and ODBE enlarge each source bit of a real tag ID into a code word by n times (N_{max} times for DBE). This indicates that $l_{tx} = C = nl$ and $|S_t| = 2^l$. Therefore, the information rate of RBE, DBE and ODBE are $\frac{1}{n}$. This indicates that RBE, DBE and ODBE incur n times larger amount of communication overhead than no encoding, and the communication efficiency decreases in inversely proportion to the value of n.

Note that although the proposed schemes sacrifice the efficiency of transmission of pseudo IDs, they improve the degree of security. Furthermore, in any encoding scheme, communication efficiency is not compatible with other desirable properties such as anonymity, error correction, etc. [10].

5.2 Expected Time for Tag Singulation

As discussed in Section 3.5, DBE and ODBE can be applied to both tree-walking and Aloha-based singulation protocols. In tree-walking singulation, the expected time for tag singulation is calculated by the number of queries that a reader requests. At first glance, the number of queries increases in proportion to the size of the set of all possible codewords, as the size of a tree becomes larger. However, surprisingly both DBE and ODBE have the same expected time for tag singulation in tree-walking singulation as the privacy masking scheme without encoding, as long as the number of tags in the system is the same.

Theorem 1 *In tree-walking singulation, DBE and ODBE have the same expected time for tag singulation as the no encoding scheme, as long as the number of tags in the system is the same.*

Proof: We prove the above claim by showing the expected number of queries that a reader requests is independent from the length of pseudo IDs. When a reader sends a query with a bit string, if more than one node have the same prefix in their IDs, collision will occur. When the length of a prefix, denoted by l_p, is one and the number of tags in a system is T, approximately half of T tags reply to the query. Given a prefix with the length l_p, responses from tags collide with the probability of $1 - (1 - \frac{1}{2})^{T-1}$. Similarly, when $l_p = 2$, the probability that collision happens is $1 - (1 - \frac{1}{4})^{T-1}$. Hence, given a prefix with the length l_p and T, the probability that responses from tags collide is calculated by $1 - (1 - \frac{1}{2^{l_p}})^{T-1}$. This indicates that the probability that responses collide is independent from the pseudo ID length. The number of queries increases in proportion to the number of collisions in responses from tags. Therefore, the number of queries is independent from the length of pseudo IDs. This completes the proof. □

In Aloha-based singulation, the expected time to identify all tags in the system is measured by the number of frames. Intuitively, encoding schemes do not affect the expected time to tag singulation, which is shown in the following Theorem.

Theorem 2 *In Aloha-based singulation, DBE and ODBE have the same expected time for tag singulation as no encoding scheme, as long as the number of tags in the system is the same.*

Proof: We prove the above claim by showing the expected number of frames that a reader sends is independent from the length of pseudo IDs. Let s (here $s \geq T$) be the number of slots in a frame, and the length of pseudo IDs is $N_{max} \cdot l$ for DBE and $n \cdot l$ for ODBE. Given a frame, each tag sends its pseudo ID in a randomly selected slot. The probability that more than one tag chooses the same slot is $1 - \frac{T!}{T^s}$. If collision occurs, collided tags select a different slot in the next frame. Denoting s_t the number of idle slot at t-th cycle, the probability that more than one tag chooses the same slot is:

$$1 - \frac{(T-(s-s_t))!}{(T-(s-s_t))^{s_t}} \tag{32}$$

Note that $T-(s-s_t)$ is the number of tags that have not being assigned slots. Equation 32 shows the probability that collision occurs is independent from the size of pseudo IDs. The expected number of frames to identify all tags in the system is dominated by frequency of collisions rather than the length of pseudo IDs. This concludes the proof. □

Note that DBE and ODBE enlarge the real tag ID, and, hence, the size of slots in frames needs to be enlarged by N_{max} times in DBE and n times in ODBE. We consider the increase of the slot size as communication overhead, since it affects the amount of data transmission. This validates our discussion in the previous section, which states that DBE and ODBE increase communication overhead by N_{max} times in DBE and n times in ODBE.

6 Experimental Validation

To evaluate the performance of the proposed DBE and ODBE schemes, we compared our techniques with no encoding [3] and RBE [18] under the environment with trusted masking device (TMD) deployment. In this section, the simulation configurations, the simulation results and the comparisons between the simulation and analytical results are presented.

6.1 Simulation Configurations

In our simulation, the length of the original ID is set to be 96 bits as defined in EPC Class1 Gen2 [8]. The value of n in RBE and ODBE, as well as N_{max} in DBE, range from 2 to 10. For the singulation protocol, the adaptive query splitting mechanism [21] is used to identify 100 tags in the RFID system. For a given configuration, 1,000 simulations are conducted. In order to achieve a fair comparison, the value of n for ODBE and RBE corresponds to N_{max} used for DBE. Consequently, the communication overhead for all protocols is exactly the same. Table 2 lists all of the simulation parameters.

TABLE 2
Simulation parameters.

Parameter	Value
Number of tags	100
Length of the original ID	96-bit
Value of n and N_{max}	2 to 10
Number of interrogation cycles	1 to 1000
Number of simulations	1000

6.2 Simulation Results

Figure 7 illustrates the degree of anonymity for generated encoded IDs as a function of the value of n. In this figure, each point indicates the average anonymity for each encoded scheme, and the range represents the span of anonymity values obtained by the simulations. As the value of n increases, the degree of anonymity increases for all protocols DBE, ODBE and RBE. Even when the value of n is small, DBE and ODBE achieve very high anonymity compared with RBE. For example, for $n = 2$, the anonymity of DBE and ODBE already reaches 0.99. These results clearly illustrate that our DBE and ODBE schemes achieve a stronger protection than RBE.

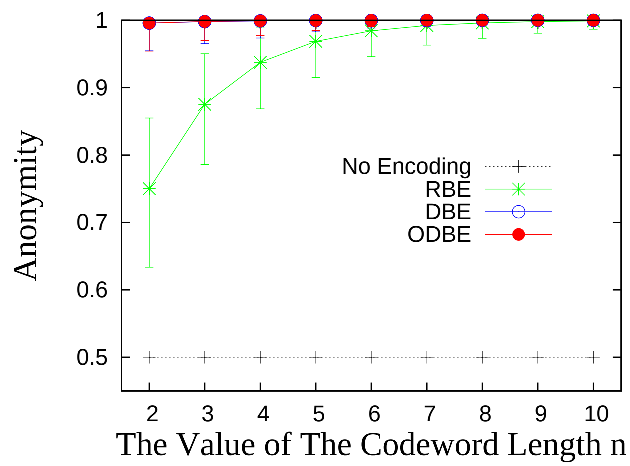

Fig. 7. Anonymity.

Figure 8 displays the time that an attacker needs to crack a tag ID as a function of the value of n. In this figure, each point depicts the average required time to crack an ID for each encoded scheme, and the range represents the extent of the required time obtained by the simulations. Attackers accumulate readable bits across interrogation cycles, and, when all bits of an encoded ID are identified, the original ID is disclosed. As we can see in Figure 8, for all protocols except the no encoding scheme (privacy masking), a longer time is required to crack an ID as the value of n increases. ODBE always requires more time than the other protocols. For DBE, when n is less than 7, it performs better than RBE. In reality, it is very unlikely for attackers to be near a tag for more than 1,000 interrogation cycles. Consequently, the results indicate that our DBE and ODBE schemes provide much stronger privacy protection for RFID systems.

Figure 9 shows the communication overhead with respect to the value of n. It is clear that for larger values of n, a tag has to transmit more bits, and the communication overhead increases proportionally. As discussed in the Dynamic Bit Encoding Scheme subsection of Section 3, theoretically the overhead of DBE and ODBE are the same as RBE and the experimental results validate this assertion. As illustrated in Figures 7 and 9, our DBE and ODBE achieve higher anonymity than RBE by paying the same overhead. Since passive tags have limited computational resources, achieving high anonymity with low overhead is very important. Our simulation results demonstrate

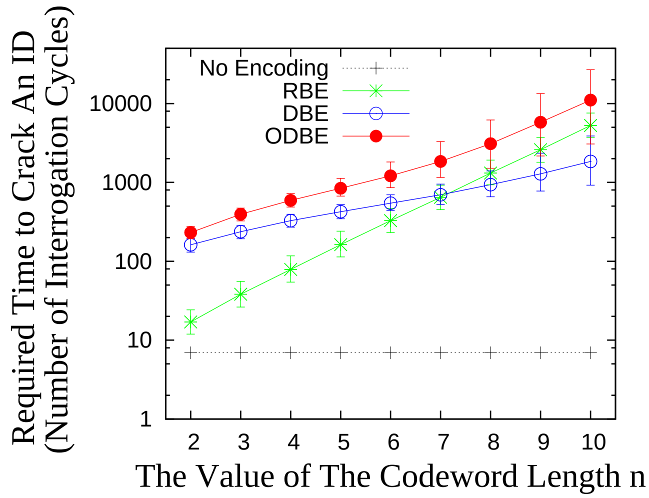

Fig. 8. Time to crack IDs.

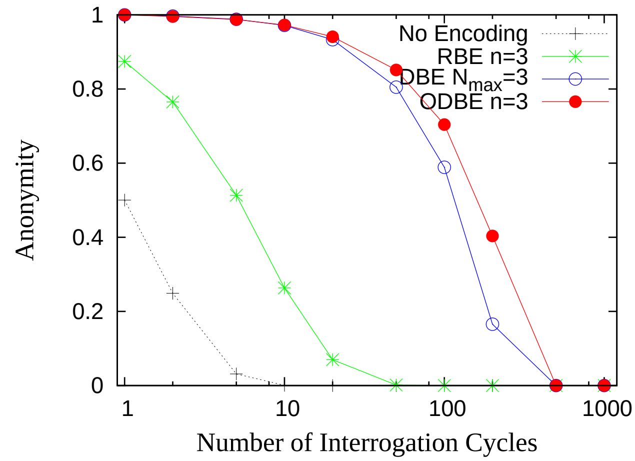

Fig. 10. Anonymity for correlation attack $n = 3$.

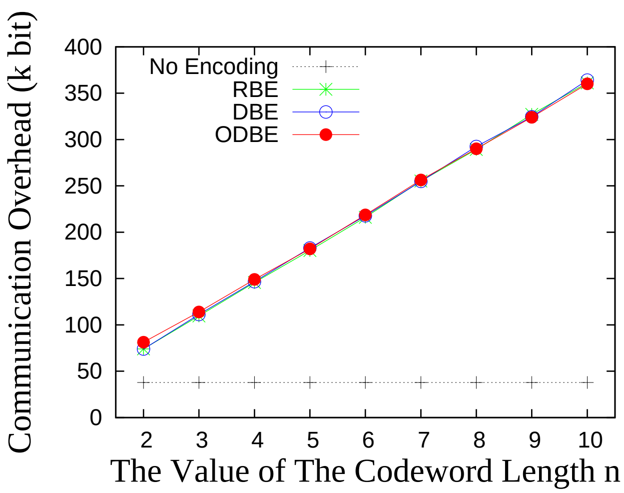

Fig. 9. Communication overhead.

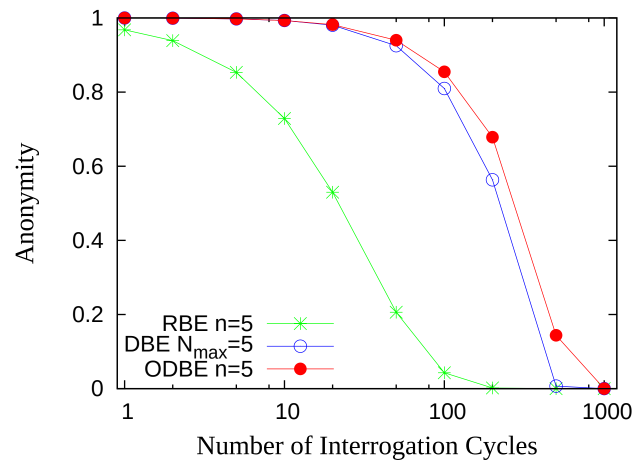

Fig. 11. Anonymity for correlation attack $n = 5$.

that DBE and ODBE are higher performing privacy protection schemes than RBE.

Figures 10, 11, and 12 demonstrate the degree of anonymity for correlation attacks for three values of n with respect to the number of interrogation cycles. We assume that each tag generates its encoded ID every interrogation cycle. As can be seen from the three figures, the larger the value of n, the higher a degree of anonymity is achieved. DBE and ODBE always accomplish a higher anonymity than the other two protocols. Also, compared with DBE, ODBE has a slightly higher anonymity. These figures suggest that DBE and ODBE are less vulnerable to correlation attacks than both RBE and the pure privacy masking scheme.

6.3 Comparison between Analytical and Simulation Results

To validate our analytical models it is important to observe a good correspondence between the analytical and simulation results.

Figure 13 illustrates the degree of anonymity with respect to the value of n. In this figure, each point symbolizes the average anonymity for each encoded scheme, and the range represents the span of anonymity values obtained by the simulations. As can be seen from the graphs, the correspondence between the analytical and simulation results for DBE and ODBE is excellent with a marginal difference of only 10^{-4}. This implies that our analytical models provide very accurate estimations in terms of anonymity for both DBE and ODBE.

Figures 14 and 15 present the degree of anonymity for DBE and ODBE against the correlation attack as a function of the number of interrogation cycles. As can be seen from these figures, there are no significant differences between analytical and simulation results.

7 Conclusion and Future Work

Privacy protection is one of the most important aspects of RFID applications. In this paper, we have proposed two bit encoding schemes for backward channel protection, namely Dynamic Bit Encoding and Optimized Dynamic Bit Encoding. In our design, the codeword length is dynamically changed for each source bit. This increases the level of difficulty for attackers to calculate and identify the original tag IDs. The simulation results show that our DBE and ODBE outperform previous solutions under conditions of original ID guessing

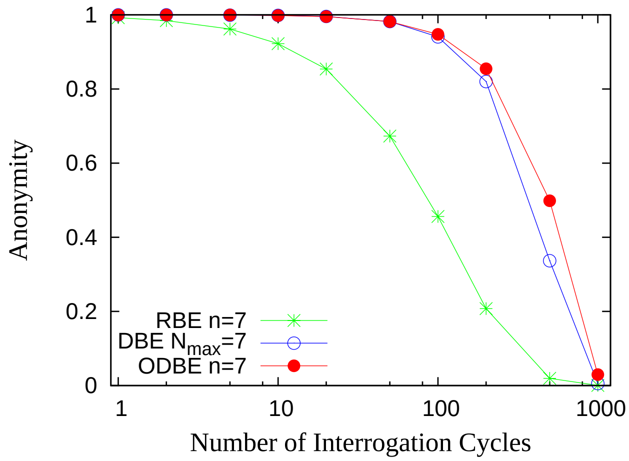

Fig. 12. Anonymity for correlation attack $n = 7$.

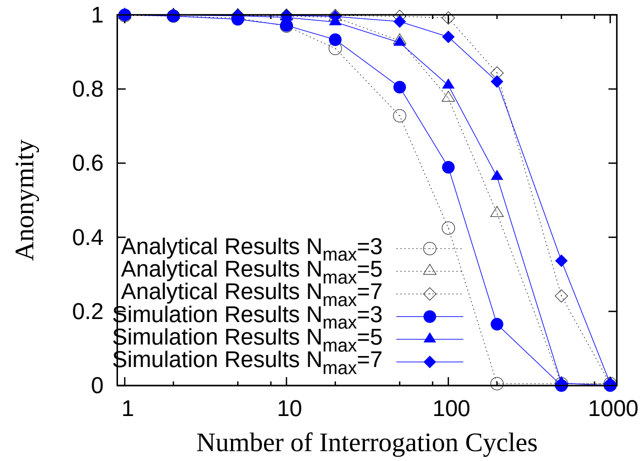

Fig. 14. Anonymity of DBE for correlation attack.

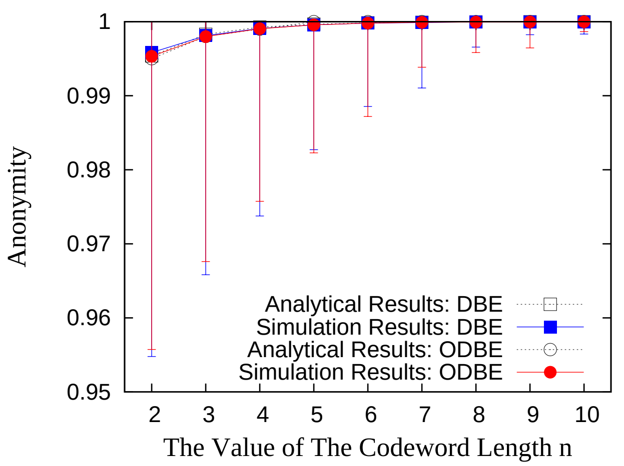

Fig. 13. Anonymity.

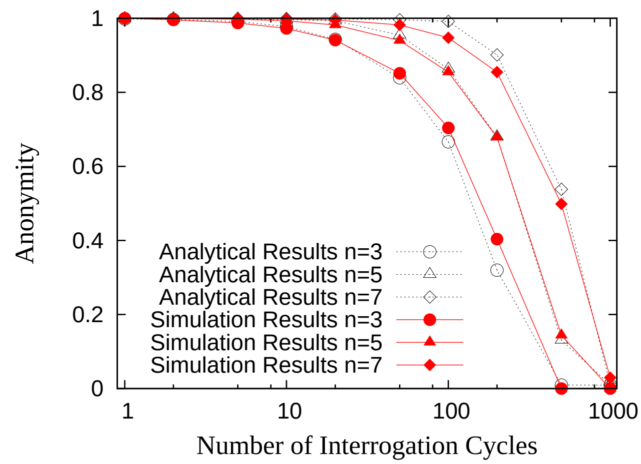

Fig. 15. Anonymity of ODBE for correlation attack.

and correlation attacks. In addition, analytical models are created to estimate the correct guess probability, the anonymity of an encoded ID and the anonymity of a tag. Furthermore, the communication overhead and time complexity of both DBE and ODBE are analyzed. The analytical models are validated through comparisons with simulation results.

While our DBE and ODBE schemes provide much stronger backward channel protection in RFID systems than existing solutions, both of them bring about communication overhead. Specifically the communication overhead increases in proportion to the codeword length. In the future, we plan to develop mechanisms which incur less communication overhead, preferably with the same backward channel protection capability.

ACKNOWLEDGEMENTS

This research has been funded in part by the National Science Foundation grants CNS-0831502 (CT) and CNS-0855251 (CRI). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

[1] Rudolf Ahlswede, Ning Cai, Shuo-Yen Robert Li, and Raymond W. Yeung. Network Information Flow. *IEEE Transactions on Information Theory*, 46(4):1204–1216, 2000.

[2] Claude Castelluccia and Gildas Avoine. Noisy Tags: A Pretty Good Key Exchange Protocol for RFID Tags. In *International Conference on Smart Card Research and Advanced Applications*, pages 289–299, 2006.

[3] Wonjoon Choi and Byeong hee Roh. Backward Channel Protection Method for RFID Security Schemes Based on Tree-Walking Algorithms. In *IEEE International Conference on Computational Science and Applications*, pages 279–287, 2006.

[4] Wonjoon Choi, Myungchul Yoon, and Byeong hee Roh. Backward Channel Protection Based on Randomized Tree-Walking Algorithm and Its Analysis for Securing RFID Tag Information and Privacy. *IEICE Transactions*, 91-B(1):172–182, 2008.

[5] Harry K. H. Chow, King Lun Choy, W. B. Lee, and K. C. Lau. Design of a RFID Case-based Resource Management System for Warehouse Operations. *Expert Syst. Appl.*, 30(4):561–576, 2006.

[6] C. Diaz, S. Seys, J. Claessens, and B. Preneel. Towards Measuring Anonymity. In *Privacy Enhancing Technologies Workshop (PET)*, 2002.

[7] Tassos Dimitriou. A Secure and Efficient RFID Protocol that could make Big Brother (partially) Obsolete. In *PerCom*, pages 269–275, 2006.

[8] EPCglobal. EPC Radio-Frequency Identity Protocols Class-1 Generation-2 UHF RFID Protocol for Communications at 860MHz-960MHz version 1.0.9.

[9] Simson L. Garfinkel, Ari Juels, and Ravikanth Pappu. RFID Privacy: An Overview of Problems and Proposed Solutions. *IEEE Security & Privacy*, 3(3):34–43, 2005.

[10] D. C. Hankerson, Gary Hoffman, D. A. Leonard, Charles C. Lindner, K. T. Phelps, C. A. Rodger, and J. R. Wall. *Coding Theory and Cryptography: The Essentials*. Marcel Dekker, 2000.

[11] Chenling Huang and Hao Min. A New Method of Synchronization for RFID Digital Receivers. In *International Conference on Solid-State and Integrated Circuit Technology (ICSICT06)*, pages 1595–1597, 2006.

[12] Rahul Jain, Jaikumar Radhakrishnan, and Pranab Sen. Privacy and Interaction in Quantum Communication Complexity and a Theorem about the Relative Entropy of Quantum States. In *IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 429–438, 2002.

[13] Ari Juels. Minimalist Cryptography for Low-Cost RFID Tags. In *International Conference on Security in Communication Networks (SCN)*, pages 149–164, 2004.

[14] Ari Juels. RFID Security and Privacy: A Research Survey. *IEEE Journal on Selected Areas in Communications*, 24(2):381–394, 2006.

[15] Ari Juels, Ravikanth Pappu, and Bryan Parno. Unidirectional Key Distribution Across Time and Space with Applications to RFID Security. In *USENIX Security Symposium*, pages 75–90, 2008.

[16] Ari Juels, Paul Syverson, and Dan Bailey. High-Power Proxies for Enhancing RFID Privacy and Utility. In *Privacy Enhancing Technologies (PET)*, 2005.

[17] Yingjiu Li and Xuhua Ding. Protecting RFID Communications in Supply Chains. In *ACM Symposium on Information, Computer and Communications Security (ASIACCS)*, pages 234–241, 2007.

[18] Tong-Lee Lim, Tieyan Li, and Sze-Ling Yeo. Randomized Bit Encoding for Stronger Backward Channel Protection in RFID Systems. In *IEEE International Conference on Pervasive Computing and Communicaitons (PerCom)*, pages 40–49, 2008.

[19] Li Lu, Jinsong Han, Lei Hu, Yunhao Liu, and Lionel M. Ni. Dynamic Key-Updating: Privacy-Preserving Authentication for RFID Systems. In *PerCom*, pages 13–22, 2007.

[20] Robert Metcalfe and David Boggs. Ethernet: Distributed Packet Switching for Local Computer Networks. *Commun. ACM*, 19(7):395–404, 1976.

[21] Jihoon Myung, Wonjun Lee, Jaideep Srivastava, and Timothy K. Shih. Tag-Splitting: Adaptive Collision Arbitration Protocols for RFID Tag Identification. *IEEE Trans. Parallel Distrib. Syst.*, 18(6):763–775, 2007.

[22] Miyako Ohkubo, Koutarou Suzuki, and Shingo Kinoshita. RFID Privacy Issues and Technical Challenges. *Commun. ACM*, 48(9):66–71, 2005.

[23] Melanie R. Rieback, Bruno Crispo, and Andrew S. Tanenbaum. Keep on Blockin' in the Free World: Personal Access Control for Low-Cost RFID Tags. In *Security Protocols Workshop*, pages 51–59, 2005.

[24] Brian L. Dos Santos and Lars S. Smith. RFID in the Supply Chain: Panacea or Pandora's Box? *Commun. ACM*, 51(10):127–131, 2008.

[25] Sanjay E. Sarma, Stephen A. Weis, and Daniel W. Engels. RFID Systems and Security and Privacy Implications. In *Cryptographic Hardware and Embedded Systems*, pages 454–469, 2002.

[26] Andrei Serjantov and George Danezis. Towards an Information Theoretic Metric for Anonymity. In *Privacy Enhancing Technologies Workshop (PET)*, 2002.

[27] Claude E Shannon and Warren Weaver. *The Mathematical Theory of Communication*. University of Illinois Press, 1998.

[28] Sarah Spiekermann and Sergei Evdokimov. Critical RFID Privacy-Enhancing Technologies. *IEEE Security & Privacy*, 7(2):56–62, 2009.

[29] Roy Want. The Magic of RFID. *ACM Queue*, 2(7):40–48, 2004.

[30] Stephen A. Weis. *Security and Privacy in Radio-Frequency Identification Devices*. Masters Thesis, Massachusetts Institute of Technology, 2005.

[31] Stephen A. Weis, Sanjay E. Sarma, Ronald L. Rivest, and Daniel W. Engels. Security and Privacy Aspects of Low-Cost Radio Frequency Identification Systems. In *Security in Pervasive Computing*, pages 201–212, 2003.

[32] Shengli Zhang, Soung Chang Liew, and Patrick P. Lam. Hot Topic: Physical-Layer Network Coding. In *MOBICOM*, pages 358–365, 2006.

APPENDIX

A. Proof of the negligible probability of DBE Pseudo ID collisions

Proof: Two DBE pseudo IDs collide with negligible probability.

Let $poly(N_{max}, l)$ be any polynomial function of the codeword length N_{max} and the ID length l. Since the first source bits of two different IDs could be the same and the first codeword length is always N_{max}, the first codewords of two different IDs could be the same. The probability of two first source bits to be the same is $\frac{1}{2}$, and the probability of first codewords to be the same is $\frac{1}{2^{N_{max}-1}}$. The rest of $(l-1)\cdot N_{max}$ bits in a pseudo ID are randomly encoded based on randomly encoded previous codewords. Therefore, the probability that two pseudo IDs collide is $\frac{1}{2}\cdot\frac{1}{2^{N_{max}-1}}\cdot\frac{1}{2^{(l-1)N_{max}}}$, which is smaller than $\frac{1}{poly(N_{max},l)}$ when N_{max} and l are sufficiently large (which can be easily fulfilled by current RFID standards). Therefore, the chance of a pseudo ID collision is negligible. This completes the proof. ∎

B. Proof of Equation 20

Proof: The proof is by induction on t. The proposition is $R(t) = l\cdot(1-\frac{1}{2^t})$.

Induction base: When $t=1$, $R(1) = l(\cdot 1-\frac{1}{2}) = l\cdot\frac{1}{2}$. The proposition is true.

Induction step: We assume $R(t) = l\cdot(1-\frac{1}{2^t})$ to prove $R(t+1) = l\cdot(1-\frac{1}{2^{t+1}})$.

$$\begin{aligned} R(t+1) &= R(t) + (l - R(t))\cdot\frac{1}{2} \\ &= \frac{1}{2}\cdot R(t) + l\cdot\frac{1}{2} \\ &= \frac{1}{2}\cdot\{l\cdot(1-\frac{1}{2^t}\} + l\cdot\frac{1}{2} \\ &= l\cdot(1-\frac{1}{2^{t+1}}) \end{aligned}$$

This concludes the proof. ∎

C. Proof of Equation 23

Proof: The proof is by induction on t. The proposition is $R(t) = l\cdot\{1-(1-\frac{1}{2^n})^t\}$.

Induction base: When $t=1$, $R(1) = l\cdot\{1-(1-\frac{1}{2^n})\} = l\cdot\frac{1}{2^n}$. The proposition is true.

Induction step: We assume $R(t) = l\cdot\{1-(1-\frac{1}{2^n})^t\}$ to prove $R(t+1) = l\cdot\{1-(1-\frac{1}{2^n})^{t+1}\}$.

$$\begin{aligned} R(t+1) &= R(t) + (l - R(t))\cdot\frac{1}{2^n} \\ &= (1-\frac{1}{2^n})\cdot R(t) + l\cdot\frac{1}{2^n} \\ &= (1-\frac{1}{2^n})\cdot l\cdot\{1-(1-\frac{1}{2^n})^t\} + l\cdot\frac{1}{2^n} \\ &= l\cdot\{1-(1-\frac{1}{2^n})^{t+1}\} \end{aligned}$$

This concludes the proof. ∎

D. Proof of Equation 29

Proof: The proof is by induction on t. The proposition is $R(t) = t \cdot \sum_{k=1}^{l} k \cdot pr_k$.

Induction base: When $t = 1$, $R(1) = \sum_{k=1}^{l} k \cdot pr_k$. The proposition is true.

Induction step: We assume $R(t) = t \cdot \sum_{k=1}^{l} k \cdot pr_k$ to prove $R(t+1) = (t+1) \cdot \sum_{k=1}^{l} k \cdot pr_k$.

$$\begin{aligned} R(t+1) &= R(t) + \sum_{k=1}^{l} k \cdot pr_k \\ &= t \cdot \sum_{k=1}^{l} k \cdot pr_k + \sum_{k=1}^{l} k \cdot pr_k \\ &= (t+1) \cdot \sum_{k=1}^{l} k \cdot pr_k \end{aligned}$$

This concludes the proof. ∎

Kazuya Sakai (S'09) received B.S. and M.S. degree in electronics engineering from Kansai University, Osaka, Japan, in 2004 and 2007, and M.S. degree in computer science from Auburn University, Auburn, AL, in 2010. Since 2010, he has been a Ph.D. student in the Department of Computer Science and Engineering at the Ohio State University. His research interests are in the area of wireless networks, mobile computing, and security. He is a student member of the IEEE.

Wei-Shinn Ku (S'02-M'07) received the Ph.D. degree in computer science from the University of Southern California (USC) in 2007. He also obtained M.S. degrees in both computer science and Electrical Engineering from USC in 2003 and 2006, respectively. He is a graduate of National Taiwan Normal University. He is currently an assistant professor with the Department of Computer Science and Software Engineering at Auburn University. His research interests include spatial and temporal data management, mobile data management, geographic information systems, and security & privacy. He has published more than 50 research papers in refereed international journals and conference proceedings. He is a member of the ACM and the IEEE. Dr. Ku can be reached at weishinn@auburn.edu.

Roger Zimmermann (S'93-M'99-SM'07) received his M.S. and Ph.D. degrees from the University of Southern California (USC) in 1994 and 1998. He is currently an Associate Professor with the Department of Computer Science at the National University of Singapore (NUS). He is also an investigator with the Interactive and Digital Media Institute at NUS. His research interests are in the areas of distributed and peer-to-peer systems, collaborative environments, streaming media architectures, geospatial database integration, and mobile location-based services. He has co-authored a book, two patents, and more than 100 conference publications, journal articles, and book chapters. He is a Senior Member of the IEEE and a member of ACM.

Min-Te Sun (S'99-M'02) received his B.S. degree in mathematics from National Taiwan University in 1991, the M.S. degree in computer science from Indiana University in 1995, and the Ph.D. degree in computer and information science from the Ohio State University in 2002. Since 2008, he has been with Department of Computer Science and Information Engineering at National Central University, Taiwan where he is currently an associate professor. His research interests include distributed algorithm design and wireless network protocol development.